FITBIR Data Submission Standards

Table of Contents

1. Introduction

The Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System was developed by the United States Department of Defense and the National Institutes of Health to share data across the entire Traumatic Brain Injury (TBI) research field and to facilitate collaboration between laboratories, as well as interconnectivity with other informatics platforms.

Sharing data, methodologies, and associated tools accelerates research progress by allowing re-analysis of data, as well as re-aggregation, integration, and rigorous comparison with other data, tools, and methods. This community-wide sharing requires

common data definitions and standards, along with comprehensive and coherent informatics approaches.

2. Who should read this document?

This document is designed to assist the researcher responsible for collecting and submitting data to FITBIR in understanding FITBIR data standards. This is typically the data manager, data coordinator, or investigator.

This document provides an explanation of FITBIR data standards to facilitate data upload for investigators.

3. Overview of the data submission process

In order to submit data to FITBIR researches are required to conform to the data submission process which includes following steps:

1. Read the FITBIR Policy.
2. Complete the Data Submission Request.
3. Scan, and sign the last "Submitter Information and Certifications" page of the Data Submission Request.
4. Email the scanned and signed "Submitter Information and Certifications" page to FITBIR or attach it to your account request. Keep the original.
5. Request a system account (unless you already have one).
6. Log in to FITBIR.
7. Create a study - this will serve as the virtual container for research data.
8. Follow the FITBIR data standards to prepare data for submission.

4. Introduction to FITBIR data submission and standards

In order to upload data to FITBIR, researchers are required to

1. Anonymize subject data and use Global Unique Identifiers (GUIDs) as subject IDs (refer to 4.2).
2. Map questions on case report forms (CRFs) to data elements in the data dictionary (refer to 4.3).
3. If the data require additional data elements and there is no match in the data dictionary, create unique data elements specific for your data (refer to 4.4, 4.5).
4. After collecting all data elements (common and unique) create electronic versions of CRFs in the system (refer to 4.6). These called form structures. A form structure creates a CSV file which should be used for data submission.
5. For each form structure, download a corresponding CSV file and populate it with data (refer to 4.6).
6. To assure the quality of data, validate CVS files with data using the Validation Tool – part of the Submission Tool module (refer to 4.7).
7. Upload valid data to FITBIR via the Upload Tool – part of the Submission Tool module (refer to 4.7).

Read more:

The Legacy Data Upload tutorial on FITBIR WIKI

4.1. Introduction to a FITBIR study

The study is a container where investigators can define and load their data. The list of studies can be accessed on FITBIR portal via Data Repository>View Studies menu.

STUDY TITLE	STUDY ID	PI	DATA TYPES	STUDY PERMISSION
Maryland MagNEtS	FITBIR-STUDY0000243	Rao Gullapalli		Read
iTAKL-youth	FITBIR-STUDY0000238	Joseph Maldjian		Read
In vivo measurement of brain biomechanics	FITBIR-STUDY0000216	Philip Bayly		Read
Frontoparietal priority maps as biomarkers for mTBI	FITBIR-STUDY0000256	Cheryl Olman, Ph.D		Read
Imaging and Biomarkers in Adolescents Cleared for Return to Play After Concussion	FITBIR-STUDY0000228	Harvey Levin		Read
Integrating Traumatic Brain Injury Model Systems Data into the Federal Interagency Traumatic Brain Injury Research Informatics System	FITBIR-STUDY0000255	Cynthia Harrrison-Felix, PhD, FACRM		Read
Late Effects of Traumatic Brain Injury (LE-TBI)	FITBIR-STUDY0000235	Wayne Gordon, PhD.		

Figure 1. The list of studies as it appears in the system

Study information

The typical study contains the following information (Figure 1):

- Title — a unique title as defined by the study Principal Investigator (PI). It is used to search for a study, select the study, view study information, etc.;
- Abstract — a short description of the study;
- Study ID — a unique study identifier assigned by the system;
- Principal Investigator — the name of the study Principal Investigator;
- Principal Investigator email — a valid email address of the study Principal Investigator;
- Data submission document — the document that explains and regulates data submissions procedures;
- Study start and end dates;
- A set of permissions that define who can access the study information and any limitations to that access;
- Other information (Figure 2).

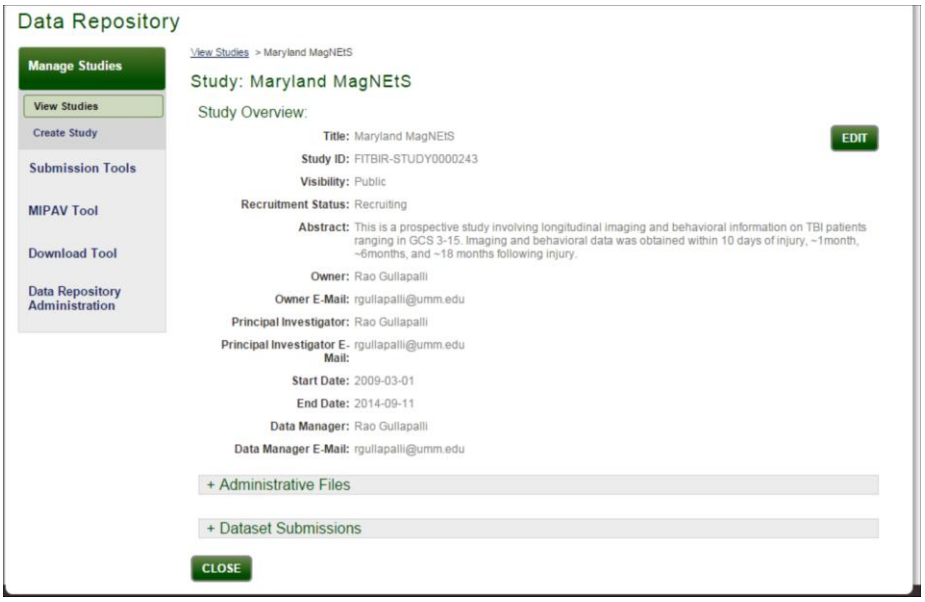

Figure 2. The study information

4.2. Introduction to data anonymization

Investigators must anonymize or de-identify the data submitted to FITBIR. Data anonymization ensures that no personally identifiable information (PII) or protected health information **(**PHI**)** could be exposed or connected to the submitted data.

In addition, de-identified data must be coded using a unique code known as a Global Unique Identifier (GUID).

GUIDs are required in order to submit data to FITBIR

The GUID is a computer-generated alphanumeric code — for example, *1A462BS* — which is unique to each research participant (i.e., each person's information or each subject's record has a different GUID). The FITBIR Operations team will teach investigators the steps required to create GUIDs.

Providing subjects' GUIDs is an essential requirement for uploading data to FITBIR. No data will be accepted to FITBIR if GUIDs are not provided.

Subject information required to generate a GUID

In order to generate a GUID for a given subject, the following PII is required:

- Complete legal given (first) name of subject at birth;
- Complete legal additional name of subject at birth, if the subject has one;
- Complete legal family (last) name of subject at birth;
- Day of birth;
- Month of birth;
- Year of birth;
- Name of city/municipality in which the subject was born;
- Country of birth.

The following optional PII can be also provided:

- Government-issued or national ID number;
- Country-issued, government-issued, or national ID;
- Physical sex of subject at birth.

How it works

Investigators collect PII from the study participants/subjects and store it in a local database not available outside of the research institution. PII is only available to a limited number of investigators. PII never goes outside of the study site.

To view data from the outside, the system uses a GUID as a subject ID. The special encryption software performs a one-way encryption (*one-way hash1*). The encrypted hash codes do not have information that could be possibly used to recreate the PII. Yet, they do have enough information to determine if a research participant's GUID already exists in the system (Figure 3).

Figure 3. Generating GUIDs. The software performs a one-way encryption

GUID process

The process for generating a GUID for a given subject is as follows:

1. The researcher logs in into FITBIR and executes the tool locally via the GUID>GUID Tool menu.
2. The researcher enters the required subject PII using the double data entry process for quality assurance (Figure 5). For **required PII**, refer to page 4.
3. In the GUID tool, PII is combined and one-way hash codes are generated.

   Note: PII cannot be extracted from these hash codes; they are strictly one-way hash algorithms**.**
4. The one-way hash codes are sent to the GUID server.
5. If the hash codes match the server's hash codes for an existing GUID, then that GUID is returned. This ensures that the same subject is assigned the same GUID across studies.
6. If the hash codes do not match, then a new random GUID is generated and returned.

No PII exposed

The GUID is a subject ID that allows researchers to share data specific to a study participant without exposing PII. The GUID is made up of random alphanumeric

characters and is NOT generated from PII/PHI. The process ensures that the GUID cannot be traced back to or identified by the original PII. In addition, the PII is never stored within the system. As such, the use of GUID has been approved by the NIH Office of General Counsel.

GUID tool

The GUID tool is a customized software application that generates a GUID for each study participant. The GUID tool is available from the FITBIR portal (account/login required).

To run the tool (Figure 4):

1. Log in into the FITBIR Portal;
2. Navigate to GUID>Create GUIDs menu and click the "Launch GUID Tool" link.

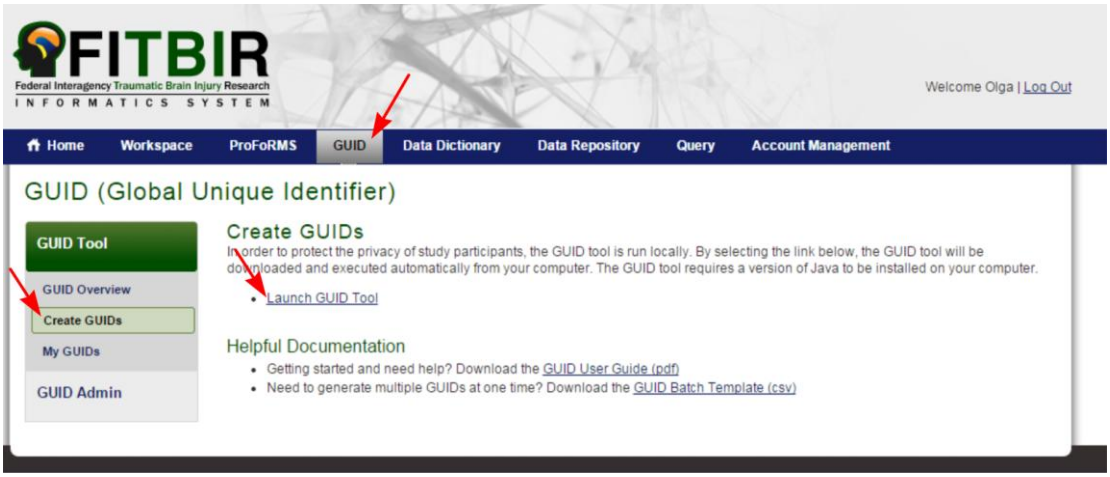

Figure 4. Running the GUID tool

Note that, in order to protect the privacy of study participants, the tool is run locally on your computer (Figure 5). The GUID tool requires the latest version of Java to be installed.

Figure 5. The GUID tool is used to create a unique subject identifier and to ensure that no PII could be possibly exposed or connected to the submitted data

GUID batch upload

If the investigators plan on generating multiple GUIDs at once, the GUID batch template functionality can be used. The batch template is a CSV file available for download from the FITBIR portal (account/login required). It can be filled out with required information and uploaded to the system using the GUID tool (Figure 6).

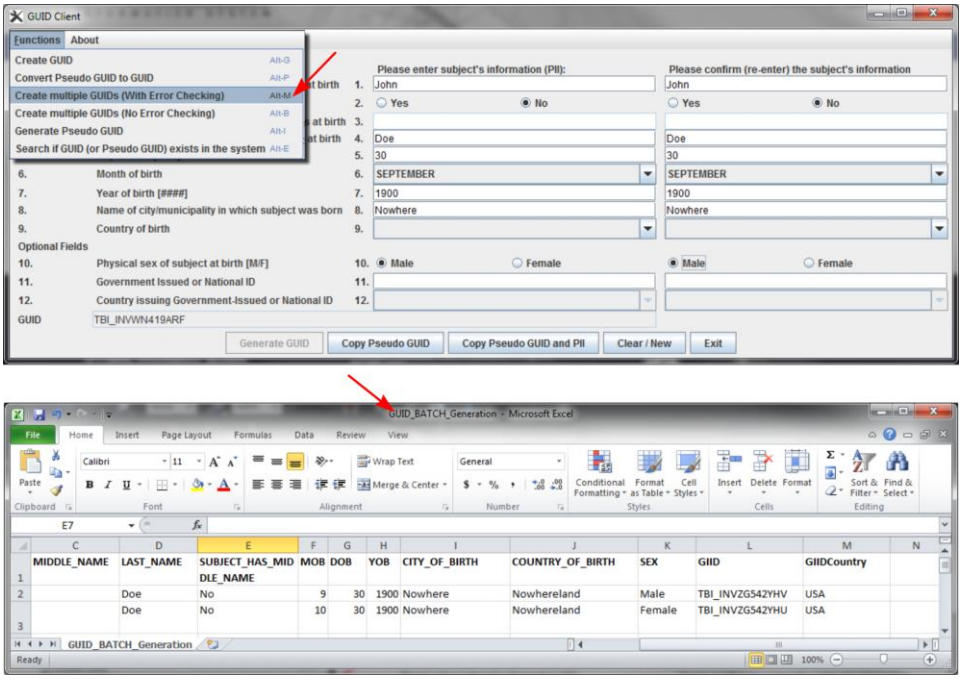

Figure 6. Generating multiple GUIDs at once using the batch generation template

Generating pseudo-GUIDs

There may be situations, such as legacy studies, where investigators do not have all the PII required to create GUIDs. In this situation, an option is available to create a pseudo GUID (Figure 7). Please, contact FITBIR Operations if considering generating pseudo GUIDs.

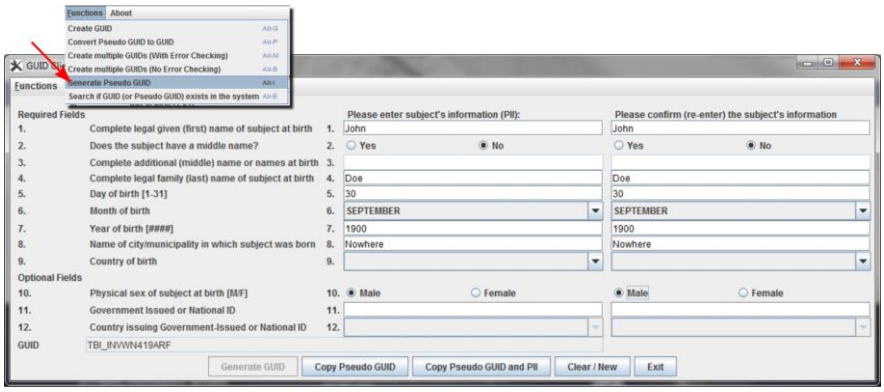

Figure 7. Creating pseudo-GUIDs

Read more:
The Generating GUIDs page on FITBIR WIKI
The GUID User Guide on FITBIR portal (only accessible when logged in)

4.3. Defining data – using existing data elements

The FITBIR data dictionary incorporates and extends the Common Data Element (CDE) definitions developed by the National Institute of Neurological Disorders and Stroke (NINDS) CDE Working Group. Importantly, the FITBIR data dictionary provides users clear and precise information about the data they are accessing.

NINDS CDE Project

The goal of the NINDS CDE Project is to develop data standards for clinical research within the neurological community. Central to this project is the creation of common definitions and data sets so that information (data) is consistently captured and recorded across studies.

CDEs are a fundamental aspect of the FITBIR data dictionary. They have been developed under the auspices of the NINDS CDE Project and its respective working groups. The purpose of the NINDS CDE Project is to standardize the collection of investigational data in order to facilitate comparison of results across studies and more effectively aggregate information into significant metadata results (Figure 8).

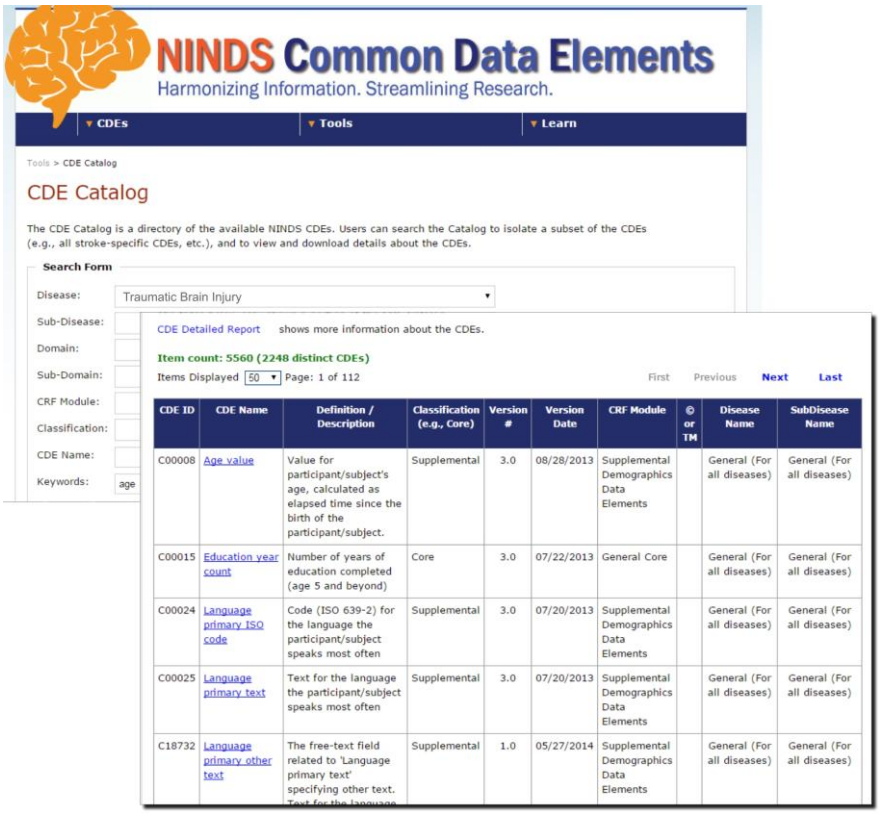

Figure 8. NINDS Common Data Elements Project

Read more:
NINDS Common Data Elements

A well-defined data dictionary

It is essential to have clearly defined and accurate information within the data dictionary, since the data that exist in FITBIR also exist in research labs and other data repositories. A well-defined data dictionary helps support the research community's vision of collaboration (Figure 9).

Investigators are expected to use the common data elements and unique data elements from the FITBIR data dictionary for collecting and submitting data.

Using CDEs to define data

To ensure the greater usability, a CDE can be used in multiple forms, as soon as it is defined within the conditions of the form (eCRF) and the form groups. The conditions of the form and its groups designate the meaning of a CDE (refer to 4.6).

What if data do not align perfectly to the CDEs?

Research data may not align perfectly to the CDEs. In this case, researchers may extend the data dictionary and create unique data elements (UDEs) specific for their study or protocol (refer to 4.4).

Note: It is possible to promote a UDE to a CDE after it is evaluated by the NINDS CDE Project's Change Control Board (CCB). The whole process might take from 6 to 12 months.

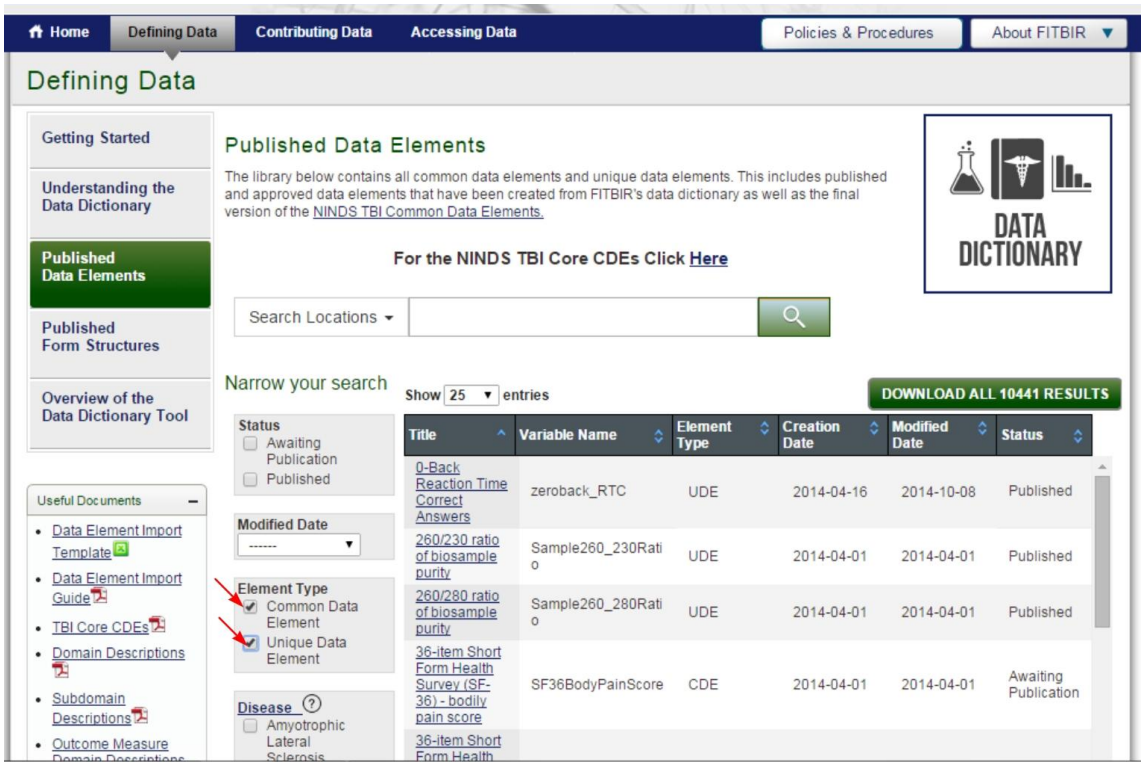

Figure 9. FITBIR data dictionary

4.4. Defining data - creating unique data elements

Before creating a new UDE, the researchers are expected to check the data dictionary and identify existing common/unique data elements to map their data. That search will assure the quality of data and integration of the whole system.

Ground rules:

- All data elements created by researchers are unique data elements (UDEs). They are study/dataset specific. UDEs must conform to the guidelines provided in the Data Element Import Guide or they will not be accepted into the FITBIR dictionary.
- Researchers must use the latest data element import template (DataElementImportTemplate.xlsx) to import their UDEs to the system.
- Before uploading to FITBIR, UDEs must be submitted to FITBIR Operations for approval. Upon approval they will be published into the data dictionary (Figure 10).

Both the Data Element Import Guide and the Data Element Import Template are available for download from the FITBIR public site (no login required).

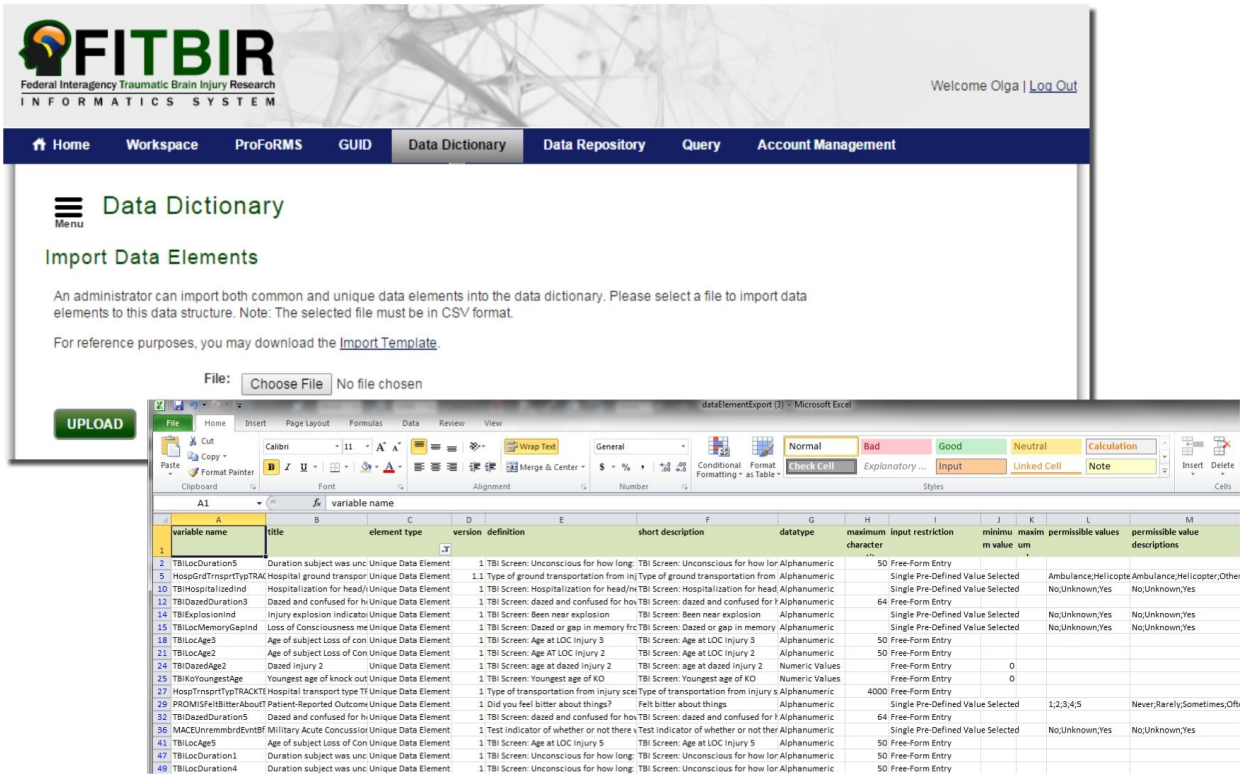

Figure 10. The data element import template is available for download from the FITBIR portal (login required)

Read more:

Published data elements on FITBIR portal

Developing Unique Data Elements on FITBIR WIKI

4.5. Defining data – permissible values

The FITBIR project uses the NINDS CDE project's policy in regards of implementing and using data elements. Therefore, FITBIR uses alphanumeric permissible values for nearly all categorical data elements (with a few exceptions).
Investigators submitting data to FITBIR are strongly encouraged to:

1. Use alphanumeric permissible values for categorical variables.
2. Use numeric permissible values only for quantitative variables where meaningful arithmetic operations can be performed on the permissible values.

Note: for FITBIR's 2.4 deployment (August 2015), the FITBIR team will deploy the ability to download coded values for categorical variables. However, please note, submission of data elements to FITBIR, which are categorical, will still need to be in alphanumeric values. For example: Gender Type must be submitted as Male or Female.

Also note that there will only be one set of mappings to coded values defined by the NINDS CDE group. Please refer to the NINDS CDE catalogue for more information about proper mappings to coded values.

4.6. Defining data – creating form structures

A form structure represents a grouping/collection of various common and unique data used to gather information for a study. A form structure is a set of CDEs and UDEs allied with the CRF, where each CDE/UDE is associated with the form's question.

Why do we require using form structures?

Investigators are expected to use standardized instruments/scales provided by FITBIR as form structures. There is much emphasis on using standardized and validated research instruments because:

- It enables comparisons of results across different studies both nationally and internationally;
- The use of standard/validated instruments increases the certainty with which the instruments accurately reflect what they are supposed to measure.

FITBIR, NINDS, and several co-sponsoring federal agencies have the common mission of developing data standards for clinical research. This includes providing form structures for standardized and validated research instruments.

If there is no form structure(s) available in the data dictionary researches can create their own form structures that correspond to standard/validated instruments.

We also need form structures in order to,

1. Upload data to the study. The form structure provides a container for uploaded data.
2. C electronic forms (eCRFs) and collect data in real time. The form structure provides a structure (or a template) for an electronic form and a container for collected data.
3. Assure the quality of uploaded data. The Data Validation tool (a part of the Submission tools module, refer to 4.7) validates the data CSV file against the

corresponding form structure and linked data elements (CDEs and UDEs). The tool then creates the submission package that can be used to submit data.

4. Guarantee that uploaded/collected data are queryable by the Query Tool.

Note: Data from all questions are expected to be submitted to FITBIR, not just summary score questions. All form structures required to have the Main group with eight data elements. Refer to page 13.

Note: FITBIR Operations, after reviewing your data collection forms, will determine if your form structure represents a standardized instrument/scale.

We strongly recommend that investigators review the published and share draft form structures available in the data dictionary, before creating their own.

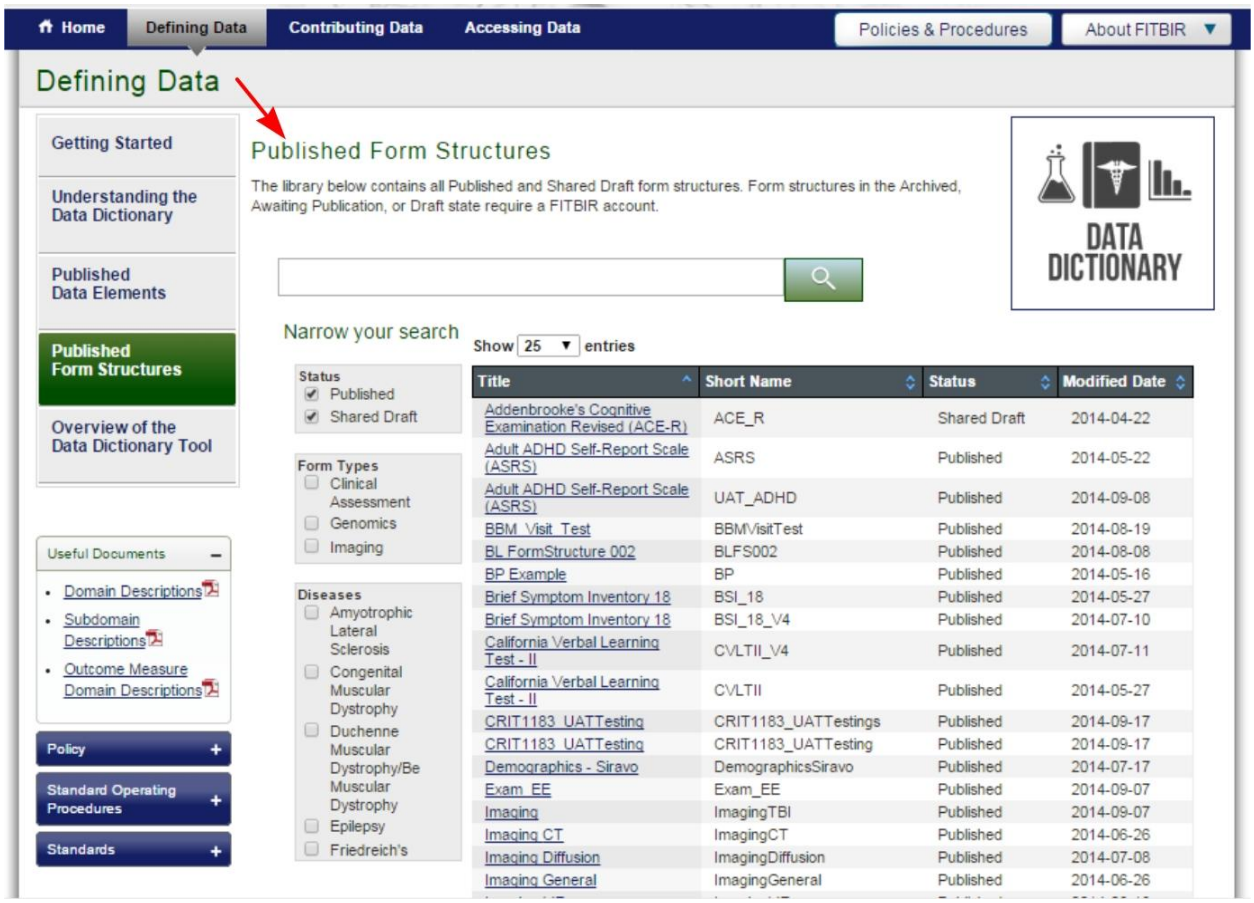

Figure 11. Form structures available via the public site (no log in needed)

Types of form structures available via the public site (fitbir.nih.gov)

Published - form structures approved by FITBIR and used by investigators to upload data; available from both FITBIR public site and portal.

Shared draft - created and supported by the NINDS CDE Project. The goal of a shared draft form structures is to provide a list of CDEs recommended for a particular CRF by the NINDS CDE Project. Investigators are required to use Shared Draft form structures as reference when create they own form structures. All shared draft form structures

require the addition of the Main and Form Administration groups before they can be published to the system.

The list of available form structures (published and shared draft) appears under the Defining Data> Published Form Structures menu on the FITBIR public site (Figure 11).

Creating form structures – the Main and Form Administration groups

Each form structure must have the Main group with eight data elements and the Form Administration group with four data elements (Table 1).

The Main and Form Administration groups contain data elements that are essential for data collection and for tracing the subject data across forms, studies, and sites.

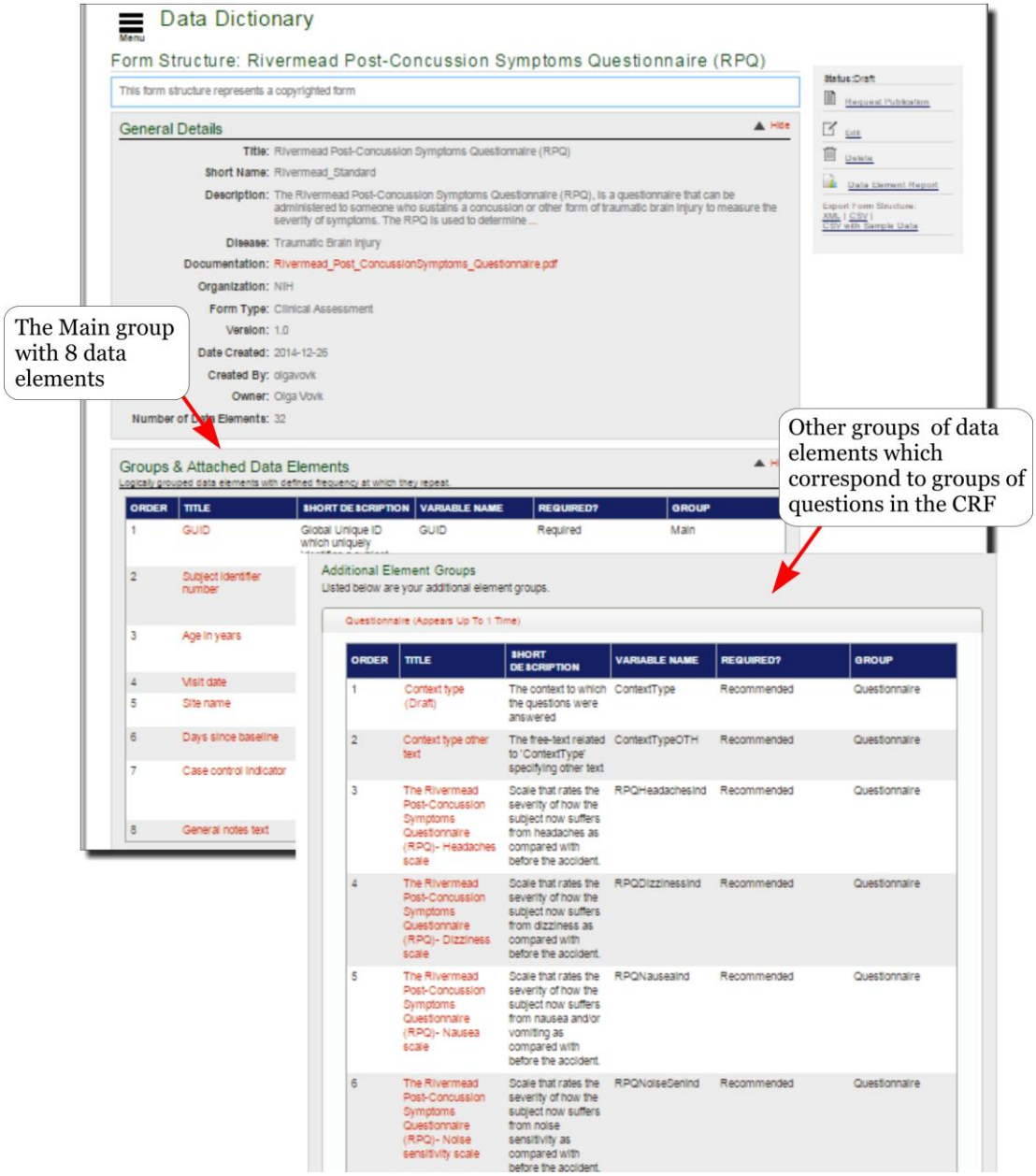

Figure 12. A form structure with the Main group of data elements as it appears in the web interface

Table 1. The data elements for the Main and Form Administration groups

Main group				
Variable Name	**Title**	**Required Type**	**Position**	**CDE/UDE**
GUID	GUID	REQUIRED	1	CDE
SubjectIDNum	Subject identifier number	OPTIONAL	2	CDE
AgeYrs	Age in years	RECOMMENDED	3	UDE
VisitDate	Visit date	RECOMMENDED	4	CDE
SiteName	Site name	RECOMMENDED	5	CDE
DaysSinceBaseline	Days since baseline	OPTIONAL	6	UDE
CaseContrlInd	Case control indicator	OPTIONAL	7	UDE
GeneralNotesTxt	General notes text	OPTIONAL	8	UDE
Form Administration group				
Variable Name	**Title**	**Required Type**	**Position**	**CDE/UDE**
ContextType	Context type	RECOMMENDED	1	UDE
ContextTypeOTH	Context type other text	RECOMMENDED	2	UDE
DataSource	Data source	RECOMMENDED	3	CDE
DataSourceOTH	Data source other text	RECOMMENDED	4	CDE

Recommendations on creating form structures

The system's web interface provides a step-by-step guide on how to create form structures, you just need to follow chevrons (Figure 13).

When creating form structures, we recommend following these steps:

1. Review the paper version of the form. Review the form layout, identify form sections.
2. Review the form structures that already exist in the system.
3. Identify form questions and related data elements.
4. Search the FITBIR data dictionary for related CDEs/UDEs.
5. If you need additional data elements, create them in the data dictionary.
6. Before creating the form structure in the system, create the MS Excel spreadsheet and gather all data elements' variable names into this spreadsheet. It will help you to review the list of data elements needed to map your data.
7. **Make sure you added the Main and Form Administration groups to the form structure**. Both groups contain data elements that are essential for data collection and for tracing the subject data across forms, studies, and sites.
8. Review the MS Excel version of the form structure and make changes if needed.
9. Log in into the system and navigate to data Dictionary>Create Form structure.
10. Use the data dictionary web interface to create the form structure in the system. If you have any questions contact the FITBIR Operations team.
11. Save the form structure. It now has the draft status, this means you can edit it and validate data against , but you cannot submit data (refer to 4.7).
12. Contact FITBIR Operations and ask them to review your form structures.
13. Upon confirmation from FITBIR Operations, validate your data against the draft form structure.

14. Save the form structure and assign permissions to use/modify the form structure to your colleagues if needed.
15. Request publication for a saved draft form structure.
16. Use the published form structure to re-validate and upload your data to the system (refer to 4.7).

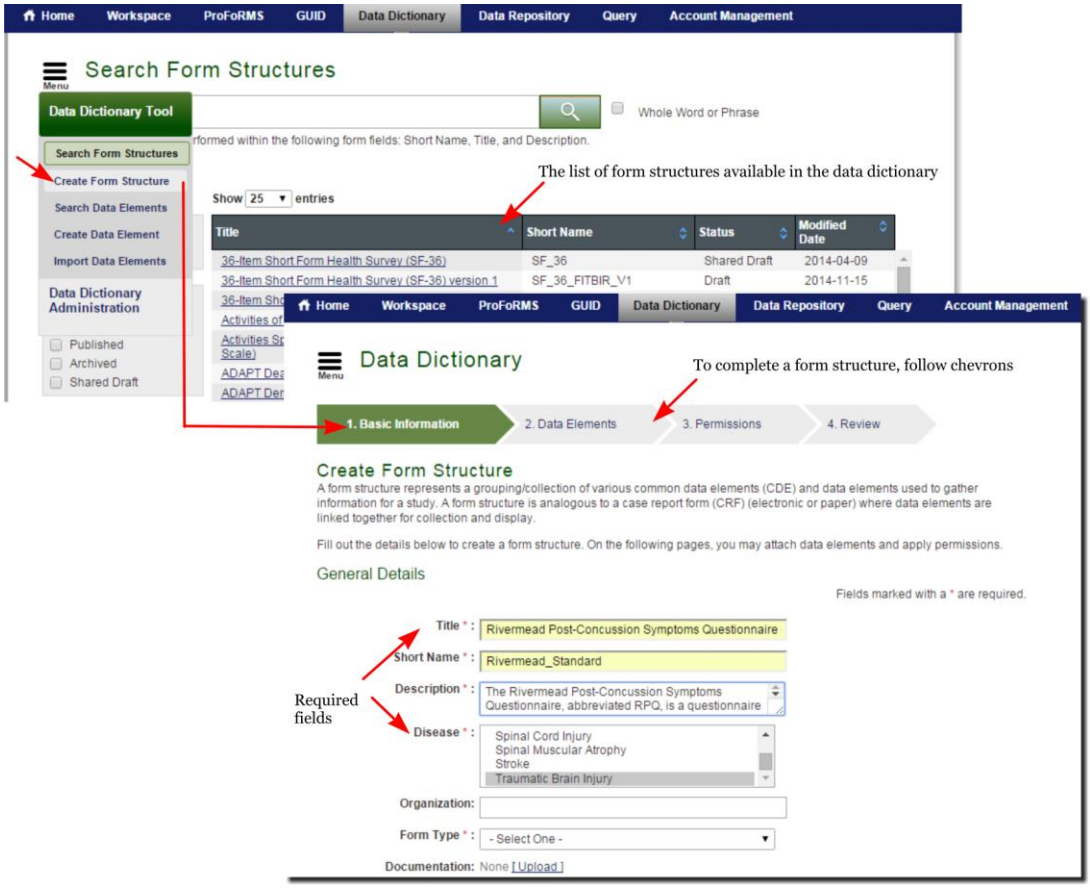

Figure 13. Creating a form structure

Read more:
The Creating Form Structures tutorial on FITBIR WIKI

4.7. Introduction to data validation and submission

The FITBIR Submission tools module is designed to provide means for data validation and upload to the system.

Validation tool

The Validation tool accepts the data as CSV files and validates the file content against the permissible values[1] defined in the data dictionary for each data element. This assures the quality of submitted data.

[1] As well as min/max values, data types, whether or not the data are required, and other essential attributes.

For those data files (CSV) that passed validation, the Validation tool creates a submission ticket and submission package (XML). The submission ticket is used by the Upload module to upload data (in the form of a corresponding submission package) to the repository. If any validation errors or warnings are found, the Validation tool provides a detailed report (Figure 14).

Recommendations: for ease of use of the Submission tools module, create an individual folder for each CSV file to upload to the database. The module will save the following information in this folder: validation error log, submission ticket and submission package.

Upload module

The Upload module accepts the submission package and submission ticket (in XML), created by the Validation tool, and uploads them to the selected study.

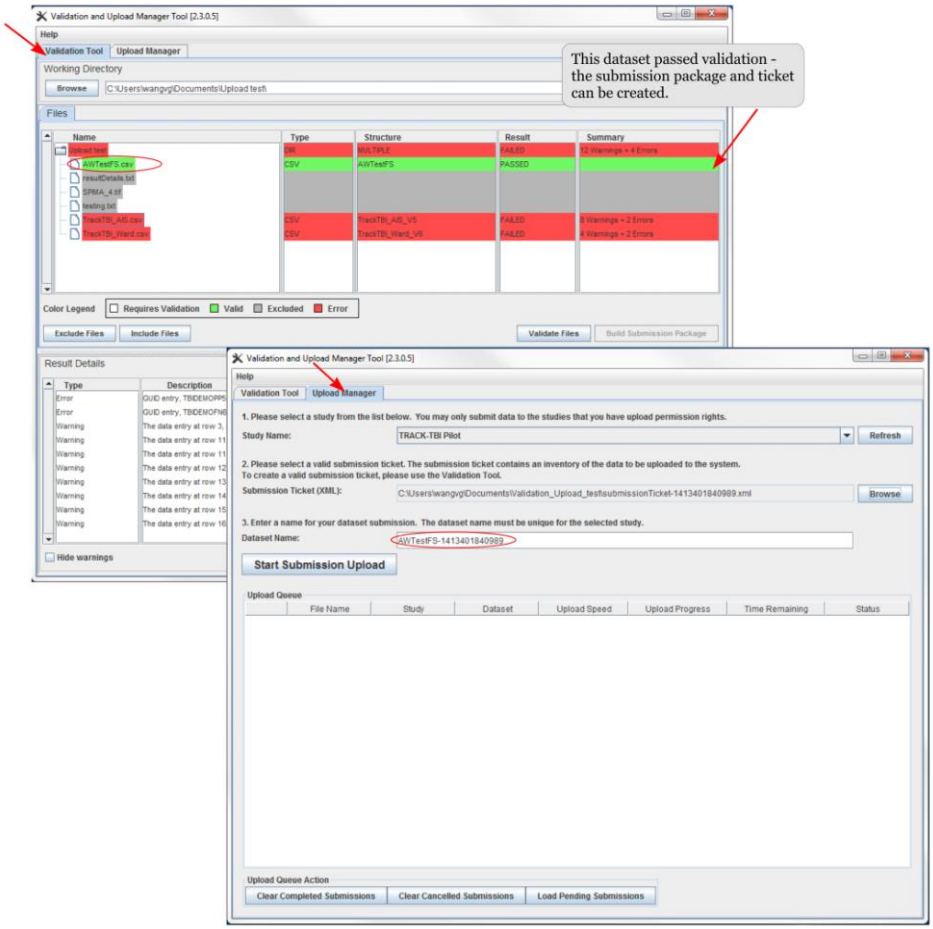

Figure 14. The FITBIR Submission tools module

Uploading imaging data

To submit imaging data to the data repository,

- Use the MIPAV tool available on the FITBIR portal to create the image submission dataset (Figure 15).
- Use the Submission tools module to re-validate the data and create a submission ticket and submission package.

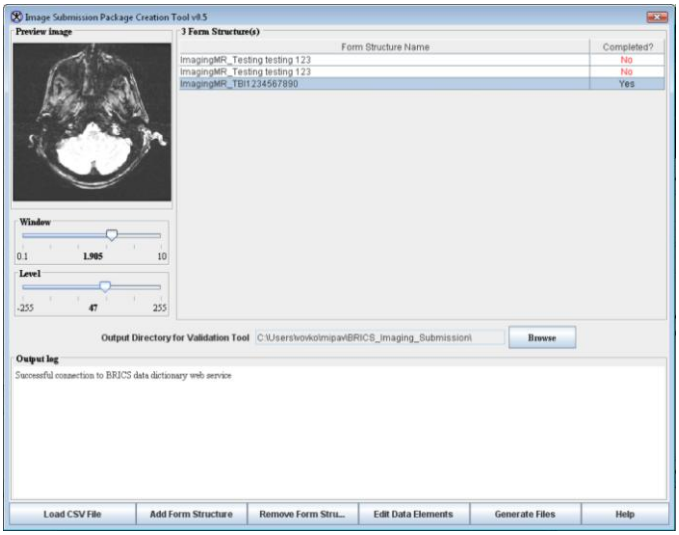

Figure 15. The MIPAV tool

Read more:

The Legacy Data Upload tutorial on FITBIR WIKI

4.8. Support provided by the FITBIR Operations team

To assist/support researchers with FITBIR data standards, the FITBIR Operations team

- Provides guidance to researchers with mapping their study variables to CDEs.
- Consults with researchers to ensure the formats of the CDEs collected are compatible with the FITBIR Informatics System.
- Assists researchers in creating UDEs (study or protocol specific), if their data point do not map to CDEs.

In addition to CDE variables, FITBIR also accepts raw data from imaging, biomarker, or physiologic studies, including:

- Study protocols;
- Manual of operations;
- Variables measured;
- Case report forms (CRFs);
- Other relevant documents.

4.9. Data submission timeline

In order to upload data on time, we recommend that investigators use the timeline shown below as a reference (Figure 16).

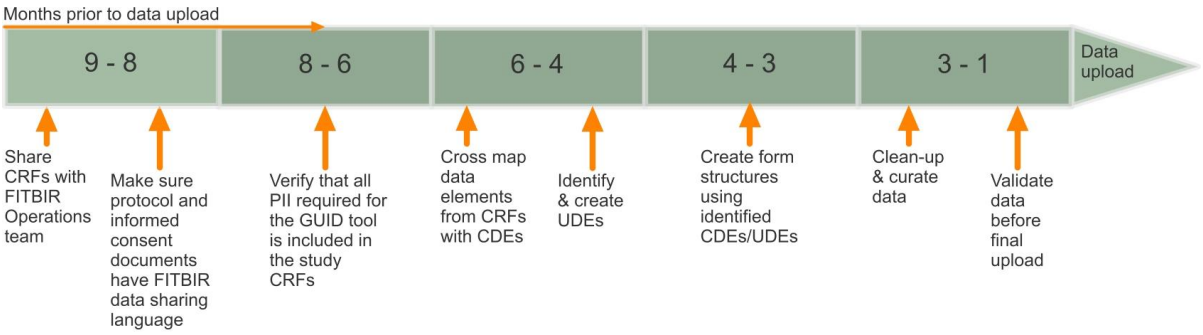

Figure 16. The recommended data submission timeline

The submission schedule for all data (except for clinical trial data) is as follows:

Data collection period	Quarterly upload due date
January 1 — March 31	June 30
April 1 — June 30	September 30
July 1 — September 30	December 31
October 1 — December 31	March 31

Note: Clinical trials are exempted from the above schedule. All data from clinical trials must be submitted within a year following the end of the performance period of the award.

4.10. Acronyms and glossary

Common data element (CDE) - a data element that can be used in multiple clinical studies, as determined by a working group of experienced clinical researchers assembled by the NINDS. The CDEs are content standards that can be applied to various data collection models and are intended to be dynamic standards that may evolve over time.

Data element - a unit of data for which the definition, identification, representation, and permissible values are specified through a set of attributes. Refer to ISO/IEC 11179 technical standard. A data element can occupy the space provided by field(s) on a paper/electronic case report form (CRF or eCRF) as well as field(s) in a database record.

eCRF - electronic Case Report Form is an electronic questionnaire specifically used in clinical research.

FITBIR - the Federal Interagency Traumatic Brain Injury Research (FITBIR) informatics system is a collaborative effort involving the NIH Institutes and Centers (ICs) and the US Army Medical Research and Material Command (USAMRMC) to develop a biomedical informatics system and data repository for Traumatic Brain Injury (TBI) research (fitbir.nih.gov). **FITBIR portal** – fitbir.nih.gov/portal/ requires log in. **FITBIR public site** – fitbir.nih.gov, no log in required.

Form structure (FS)- represents a grouping/collection of data elements used in the BRICS data dictionary. A form structure is analogous to a case report form (CRF) (electronic or paper) where data elements are linked together for collection and display.

Form structure status – represent the status of the form structure in the FITBIR data dictionary.

Status	Definition	Available	Data Validation	Data Submission
Draft	A status assigned to a new form structure that has been just created.	FITBIR portal, only for admin and owner	Yes	No
Awaiting publication	A status assigned to a form structure that was requested to publish. The previous status for this form structure was draft.	FITBIR portal for all users	Yes	No
Published	A status assigned to a form structure, which is finalized. The previous status for this form structure was awaiting publication.	FITBIR public site and portal, all users	Yes	Yes
Shared draft	Created and supported by the NINDS CDE Project. Investigators are required to use Shared Draft form structures as reference when create they own form structures.	FITBIR public site and portal, all users	No	No
Archived	A status assigned to a form structure which is no longer in use.	FITBIR portal, admin only	No	No

GUID - a Global Unique Identifier is a subject identifier that allows researchers to share data specific to a study participant without exposing personally identifiable information (PII). A GUID is made up of random alpha-numeric characters and is NOT generated from PII/PHI.

MIPAV - the Medical Image Processing, Analysis, and Visualization application enables quantitative analysis and visualization of medical images of numerous modalities such as PET, MRI, CT, or microscopy

NIH - the National Institutes of Health a part of the U.S. Department of Health and Human Services , is the nation's medical research agency—making important discoveries that improve health and save lives.

NINDS – the mission of the National Institute of Neurological Disorders and Stroke is to seek fundamental knowledge about the brain and nervous system and to use that knowledge to reduce the burden of neurological disease.

NINDS CDE Project - the goal of the National Institute of Neurological Disorders and Stroke (NINDS) CDE Project specifically is to develop data standards for clinical research within the neurological community. Central to this project is the creation of common definitions and data sets so that information (data) is consistently captured and recorded across studies.

PHI - Protected health information (PHI) is any information about health status, provision of health care, or payment for health care that can be linked to a specific individual.

PII - Personally Identifiable Information. Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual‘s identity, such as name, social security number, date and place of birth, mother‘s maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information. Refer to NIST Special Publication 800-122.

Pseudo-GUID – an Identifier created by the system when not all information required to create a GUID is available.

Study - is a container where investigators can define and load their data. The list of studies can be accessed on FITBIR portal via the Data Repository>View Studies menu.

Submission tools - the Submission Tools module assist researchers with the validation and upload of data into the repository. The validation component verifies that submitted data conforms to the required format and range values defined in the data dictionary.

Unique data element(s) - belong to a particular dataset (e.g. disease specific, form specific, or media/modality specific) and cannot be used in any other datasets. They are defined within a particular dataset to capture very specific data and do not have a life outside of that dataset, or outside of a very specific form (eCRF).